

B21

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|----|--|
| (51) International Patent Classification 6 : C07H 21/04, C12Q 1/68, C12P 19/34 | A1 | (11) International Publication Number: WO 96/40723 (43) International Publication Date: 19 December 1996 (19.12.96) |
|---|----|--|

(21) International Application Number: PCT/US96/09358
(22) International Filing Date: 4 June 1996 (04.06.96)
(30) Priority Data:
08/487,867 7 June 1995 (07.06.95) US

(81) Designated States: AU, CA, FI, IL, JP, KR, MX, NO, NZ,
European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB,
GR, IE, IT, LU, MC, NL, PT, SE).

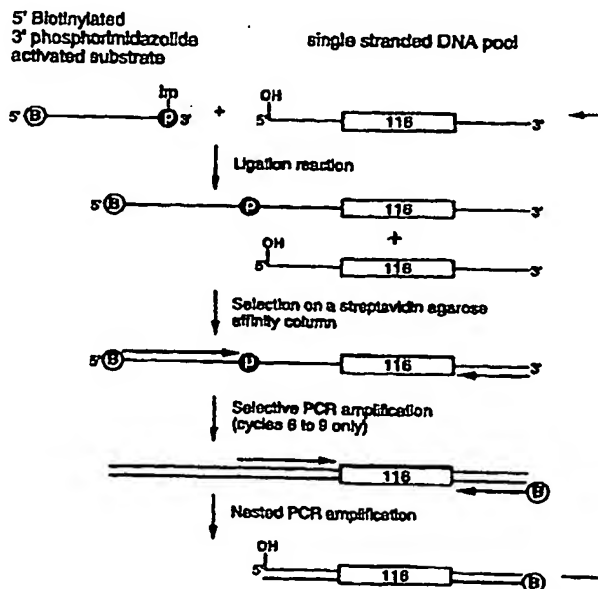
Published
With international search report.

(71) Applicant: THE GENERAL HOSPITAL CORPORATION
[US/US]; 55 Fruit Street, Boston, MA 02114 (US).

(72) Inventors: SZOSTAK, Jack, W.; 308 Commonwealth Avenue,
Boston, MA 02115 (US). CUENOUD, Bernard; Mittlere
Strasse 7, CH-4056 Basle (CH). HUIZENGA, David, E.;
Apartment 3R, 38 Forrest Street, Winthrop, MA 02152 (US).

(74) Agent: LECH, Karen, F.; Fish & Richardson P.C., 225 Franklin
Street, Boston, MA 02110-2804 (US).

(54) Title: CATALYTIC DNA



(57) Abstract

The invention features nucleic acid molecules and, in particular, DNA molecules having catalytic activity, as well as methods for obtaining and using such nucleic acid molecules.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AM | Armenia | GB | United Kingdom | MW | Malawi |
| AT | Austria | GE | Georgia | MX | Mexico |
| AU | Australia | GN | Guinea | NE | Niger |
| BB | Barbados | GR | Greece | NL | Netherlands |
| BE | Belgium | HU | Hungary | NO | Norway |
| BF | Burkina Faso | IE | Ireland | NZ | New Zealand |
| BG | Bulgaria | IT | Italy | PL | Poland |
| BJ | Benin | JP | Japan | PT | Portugal |
| BR | Brazil | KE | Kenya | RO | Romania |
| BY | Belarus | KG | Kyrgyzstan | RU | Russian Federation |
| CA | Canada | KP | Democratic People's Republic of Korea | SD | Sudan |
| CF | Central African Republic | KR | Republic of Korea | SE | Sweden |
| CG | Congo | KZ | Kazakhstan | SG | Singapore |
| CH | Switzerland | LI | Liechtenstein | SI | Slovenia |
| CI | Côte d'Ivoire | LK | Sri Lanka | SK | Slovakia |
| CM | Cameroon | LR | Liberia | SN | Senegal |
| CN | China | LT | Lithuania | SZ | Swaziland |
| CS | Czechoslovakia | LU | Luxembourg | TD | Chad |
| CZ | Czech Republic | LV | Latvia | TG | Togo |
| DE | Germany | MC | Monaco | TJ | Tajikistan |
| DK | Denmark | MD | Republic of Moldova | TT | Trinidad and Tobago |
| EE | Estonia | MG | Madagascar | UA | Ukraine |
| ES | Spain | ML | Mali | UG | Uganda |
| FI | Finland | MN | Mongolia | US | United States of America |
| FR | France | MR | Mauritania | UZ | Uzbekistan |
| GA | Gabon | | | VN | Viet Nam |

- 1 -

CATALYTIC DNABackground of the Invention

This invention relates to DNA molecules having
5 catalytic activity and methods of obtaining and using
such DNA molecules.

Ribozymes are highly structured RNA molecules that
carry out specific chemical reactions (e.g., cleavage of
RNA, cleavage of DNA, polymerization of RNA, and
10 replication of RNA), often with kinetic efficiencies
comparable to those of most engineered enzymes.

Summary of the Invention

The invention features nucleic acid molecules
having catalytic activity, as well as methods for
15 obtaining and using such nucleic acid molecules.

The methods of the invention entail sequential
in vitro selection and isolation of nucleic acid
molecules having the desired properties (e.g., catalytic
activity, such as ligase activity) from pools of single-
20 stranded nucleic acid molecules (e.g., DNA, RNA, or
modifications or combinations thereof) containing random
sequences. The isolated nucleic acid molecules are then
amplified by using, e.g., the polymerase chain reaction
(PCR).

25 The rounds of selection and amplification may be
repeated one or more times, after each round, the pool of
molecules being enriched for those molecules having the
desired activity. Although the number of desired
molecules in the initial pool may be exceedingly small,
30 the sequential selection scheme overcomes this problem by
repeatedly enriching for the desired molecules.

The pool of single-stranded nucleic acid molecules
employed in the invention may be referred to as "random
nucleic acid molecules" or as containing "random

- 2 -

sequences." These general terms are used to describe molecules or sequences which have one or more regions of "fully random sequence." In a fully random sequence, there is an approximately equal probability of A, T/U, C, or G being present at each position in the sequence. Of course, the limitations of some methods used to create nucleic acid molecules make it rather difficult to synthesize fully random sequences in which the probability of each nucleotide occurring at each position is absolutely equal. Accordingly, sequences in which the probabilities are roughly equal are considered fully random sequences.

In "partially random sequences" and "partially randomized sequences," rather than there being a 25% chance of A, T/U, C, or G being present at each position, there are unequal probabilities. For example, in a partially random sequence, there may be a 70% chance of A being present at a given position and a 10% chance of each of T/U, C, or G being present at that position. Further, the probabilities can be the same or different at each position within the partially randomized region. Thus, a partially random sequence may include one or more positions at which the sequence is fully random, one or more positions at which the sequence is partially random, and/or one or more positions at which the sequence is defined.

Partially random sequences are particularly useful when one wishes to make variants of a known sequence. For example, if one knows that a particular 50 nucleotide sequence possesses a desired catalytic activity and that positions 5, 7, 8, and 9 are critical for this activity, one could prepare a partially random version of the 50 nucleotide sequence in which the bases at positions 5, 7, 8, and 9 are the same as in the catalytically active sequence, and the other positions are fully randomized.

- 3 -

Alternatively, one could prepare a partially random sequence in which positions 5, 7, 8, and 9 are partially randomized, but with a strong bias towards the bases found at each position in the original molecule, with all of the other positions being fully randomized. This type of partially random sequence is desirable in pools of molecules from which catalytic nucleic acids are being selected. The sequence of any randomized region may be further randomized by mutagenesis during one or more amplification steps.

In addition to random or partially random sequences, it may also be desirable to have one or more regions of "defined sequence." A defined sequence is a sequence selected or known by the creator of the molecule. Defined sequence regions are useful for isolating or PCR amplifying the nucleic acid molecule because they may be recognized by defined complementary primers. The defined sequence regions may flank the random regions or be intermingled with the random regions. The defined regions can be of any length desired and are readily designed using knowledge in the art (see, for example, Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing, New York, New York (1994); Ehrlich, PCR Technology, Stockton Press, New York, New York (1989); and Innis et al., PCR Protocols, A Guide to Methods and Applications, Academic Press, Inc., San Diego, CA (1990)).

The selection method of the invention involves contacting a pool of nucleic acid molecules containing random sequences with the substrate for the desired catalytic activity under conditions (including, e.g., nucleic acid molecule concentrations, temperature, pH, and salt) which are favorable for the catalytic activity. Nucleic acid molecules having the catalytic activity are partitioned from those which do not, and the partitioned

- 4 -

nucleic acid molecules having the catalytic activity then are amplified using, e.g., PCR.

The steps of contacting, partitioning, and amplifying may be repeated any desired number of times. 5 Several cycles of selection (contacting, partitioning, and amplifying) may be desirable because after each round the pool is more enriched for the desired catalytic nucleic acids. One may choose to perform so many cycles of selection that no substantial improvement in catalytic 10 activity is observed upon further selection, or one may carry out far fewer cycles of selection.

Methods known in the art may be used at particular steps of this selection and isolation procedure, and one skilled in the art is referred to Ellington and Szostak, 15 Nature 346:818-822, 1990; Lorsch and Szostak, Nature 371:31-36, 1994; Tuerk and Gold, Science 249:505-510, 1990; and methods described herein.

In addition, one may mutagenize isolated catalytic nucleic acids in order to generate and subsequently 20 isolate molecules exhibiting improved catalytic activity. For example, one may prepare degenerate pools of single-stranded nucleic acids based on a particular catalytic nucleic acid sequence, or one may first identify important regions in a catalytic nucleic acid sequence 25 (for example, by standard deletion analysis), and then prepare pools of candidate catalytic nucleic acid molecules that include degenerate sequences at those important regions.

Those skilled in the art can readily identify 30 catalytic nucleic acid consensus sequences by sequencing a number of catalytic nucleic acid molecules and comparing their sequences. In some cases, such sequencing and comparison will reveal the presence of a number of different conserved sequences. In these 35 circumstances, one may identify a core sequence which is

- 5 -

common to most or all of the isolated sequences. This core sequence, or variants thereof, may be used as the starting point for the selection of improved catalysts. By "variant" of a sequence is meant a sequence created by
5 partially randomizing the sequence.

The size of the randomized regions employed should be adequate to provide a catalytic site. Thus, the randomized region used in the initial selection preferably includes between 10 and 300 nucleotides, for
10 example, between 25 and 180 nucleotides.

It may be desirable to increase the stringency of a selection step in order to isolate more molecules. The stringency of the selection step may be increased by decreasing substrate concentration. The stringency of
15 the catalysis selection step can be increased by decreasing the ligand concentration or the reaction time.

In one aspect, therefore, the invention features a method for obtaining a nucleic acid molecule having ligase activity. In the first step of this method, a
20 population of candidate nucleic acid molecules, each having a region of random sequence, is contacted with a substrate nucleic acid molecule and an external template. The external template is complementary to a portion of the 3' region of the substrate nucleic acid molecule and
25 a portion of the 5' region of each of the candidate nucleic acid molecules in the population. Alternatively, the external template may be complementary to a portion of the 5' region of the substrate nucleic acid molecule and a portion of the 3' region of each of the candidate
30 nucleic acid molecules in the population. Binding of the external template to the substrate nucleic acid molecule and a candidate nucleic acid molecule from the population juxtaposes the 3' region of one of the molecules with the 5' region of the other.

- 6 -

One of the terminal nucleotides (either the 5' or the 3' nucleotide) of the juxtaposed regions may contain an activated group. Activated groups that may be used in the method of the invention include, but are not limited to, 5'-phosphoro(2-methyl)imidazolide, a 5'-phosphorimidizolide, cyanogen bromide, and carbodiimides (e.g., 1-ethyl-3-(3'-dimethylaminopropyl) carbodiimide (CDI), 1-cyclohexyl-3-(2-morpholinyl-(4)-ethyl)-carbodiimide metho-p-toluenesulfonate, CDI-1, and CDI-2).

As a specific example, the activated group is a 3'-phosphorimidazolide on the 3' terminal nucleotide of the substrate. Activating groups are added to the nucleic acid molecules used in the methods of the invention by using methods known in the art.

Alternatively, if desired, this first step external templating may be omitted. It is not essential to the selection method of the invention.

In the second step of this method of the invention, a subpopulation of nucleic acid molecules having ligase activity is isolated from the population. This may be accomplished by, e.g., affinity chromatography followed by selective PCR amplification. For example, the substrate nucleic acid and/or the nucleic acid from the population may contain the first member of a specific binding pair (e.g., biotin). As a specific example, the terminal nucleotide of the substrate nucleic acid (e.g., the 5' terminal nucleotide of the substrate nucleic acid) and/or the nucleic acid molecule from the population that is not juxtaposed by the external template may be labeled with biotin. Isolation of molecules containing biotin may be accomplished by contacting the molecules with immobilized avidin, e.g., a streptavidin agarose affinity column. Other specific binding pairs known to one skilled in the art may be used in the method of the invention.

- 7 -

The isolated subpopulation may be amplified in vitro using, e.g., PCR. In selective PCR, the first primer is complementary to a sequence of the substrate nucleic acid molecule and the second primer is

5 complementary to the opposite strand of a sequence in the population. Use of these primers therefore results in amplification of only those nucleic molecules which are a product of the ligation of the substrate to a nucleic acid molecule from the population. In order to generate

10 a population of nucleic acid molecules for further rounds of selection, nested PCR amplification may be carried out using primers which preferably include the terminal nucleotides of the nucleic acid from the population that was ligated to the substrate nucleic acid.

15 The above-described steps of contacting, isolating, and amplifying may be repeated on the subpopulations of nucleic acid molecules obtained. The additional rounds of selection may be carried out in the presence or absence of the external template. Nucleic

20 acid molecules isolated using the above-described method may be subcloned into a vector (e.g., a plasmid) and further characterized by, e.g., sequence analysis.

In a second aspect, the invention features a DNA molecule capable of acting as a catalyst. A catalyst is

25 a molecule which enables a chemical reaction to proceed under different conditions (e.g., at a lower temperature, with lower reactant concentrations, or with increased kinetics) than otherwise possible.

In a third aspect, the invention features a DNA

30 molecule capable of acting as a catalyst on a nucleic acid substrate. This catalysis does not require the presence of a ribonucleotide in the nucleic acid substrate.

In a fourth aspect, the invention features a

35 nucleic acid molecule having ligase activity, e.g., DNA

- 8 -

or RNA ligase activity. The nucleic acid molecule may be DNA, RNA, or combinations or modifications thereof.

In a fifth aspect, the invention features a nucleic acid molecule capable of ligating a first
5 substrate nucleic acid to a second substrate nucleic acid. The rate of ligation catalyzed by the nucleic acid molecule of the invention is greater than the rate of ligation of the substrate nucleic acids by templating under the same reaction conditions which include such
10 variables as, e.g., substrate concentration, template/enzyme concentration, nature and quantity of base-pairing interactions between substrates and template/enzyme, type of activating group, salt, pH, and temperature. Templating is the joining of two substrate
15 nucleic acid molecules when hybridized to contiguous regions of a "template" nucleic acid strand.

In a sixth aspect, the invention features a catalytic DNA molecule capable of ligating a first
20 substrate nucleic acid to a second substrate nucleic acid. The first substrate nucleic acid contains the sequence 3'-S¹-S²-5', the second substrate nucleic acid contains the sequence 3'-S³-S⁴-5', and the catalytic DNA molecule contains the sequence 5'-E¹-TTT-E²-AGA-E³-E⁴-E⁵-E⁶-3'.

25 For these substrate and catalytic DNA molecules, S¹ contains at least two (for example, 2-100, 4-16, or 8-12) nucleotides positioned adjacent to the 3' end of S². The S¹ nucleotides are complementary to an equivalent number of nucleotides in E¹ that are positioned adjacent
30 to the 5' end of TTT.

S² contains one - three (for example, 1) nucleotides, S³ contains one - six (for example, 3) nucleotides, and the 5' terminal nucleotide of S² and the 3' terminal nucleotide of S³ alternatively contain an
35 activated group or a hydroxyl group.

- 9 -

S⁴ contains at least two (for example, 2-100, 4-16, or 8-12) nucleotides positioned adjacent to the 5' end of S³. The S⁴ nucleotides are complementary to an equivalent number of nucleotides in E⁶ that are
5 positioned adjacent to the 3' end of E⁵.

E¹ contains at least two (for example, 2-100, 4-16, or 8-12) nucleotides positioned adjacent to the 5' end of TTT. The E¹ nucleotides are complementary to an equivalent number of nucleotides in S¹ that are
10 positioned adjacent to the 3' end of S².

E² contains 0-12 nucleotides, for example, 3-4 nucleotides.

E³ contains at least two (for example, 2-100, 3-50, 5-20, or 5) nucleotides positioned adjacent to the 3' end of said AGA, said E³ nucleotides being complementary to an equivalent number of nucleotides in E⁵ that are
15 positioned adjacent to the 5' end of E⁶.

E⁴ contains at least 3 nucleotides (for example, 3-200, 3-30, 3-8, 4-6, or 5) nucleotides. Alternatively,
20 E⁴ may contain zero nucleotides. In this case, the 3' end of E³ and the 5' end of E⁵ would not be linked to another nucleic acid segment (e.g., E⁴), and the enzyme therefore would be made up of two separate nucleic acid molecules (the first containing 5'-E¹-TTT-E²-AGA-E³-3',
25 and the second containing 5'-E⁵-E⁶-3').

E⁵ contains at least two (for example, 2-100, 3-50, 5-20, or 5) nucleotides positioned adjacent to the 5' end of E⁶. The E⁵ nucleotides are complementary to an equivalent number of nucleotides in E³ that are
30 positioned adjacent to the 3' end of AGA.

E⁶ contains at least two (for example, 2-100, 4-16, or 8-12) nucleotides positioned adjacent to the 3' end of E⁵. The E⁶ nucleotides are complementary to an equivalent number of nucleotides in S⁴ that are
35 positioned adjacent to the 5' end of S³.

- 10 -

In the case of long stem structures formed by, e.g., S^1 and E^1 , S^4 and E^6 , or E^3 and E^5 , the stem structures may contain mismatches, provided that a stem structure is maintained.

5 The 5' most nucleotide of S^2 , the 3' most nucleotide of S^3 , and the second 3' most nucleotide of S^3 may be complementary to the 5' most nucleotide of E^2 , the second 5' most nucleotide of E^2 , and the third 5' most nucleotide of E^2 , respectively. In addition, E^2 may
10 contain four nucleotides, and the third 3' most nucleotide of S^3 may be complementary to the fourth 5' most nucleotide of E^2 .

In a seventh aspect, the invention features a method of ligating a first nucleic acid molecule to a
15 second nucleic acid molecule. In this method, the first and second nucleic acid molecules are contacted with a nucleic acid molecule having ligase activity (e.g., DNA ligase activity). The nucleic acid molecule having
ligase activity, as well as the first and second nucleic
20 acid molecules may contain DNA, RNA, or modifications or combinations thereof

The ease with which DNA oligonucleotides can be synthesized and their relatively high stability represent major advantages over other biopolymer catalysts, such as
25 proteins and RNA, for, e.g., industrial, research, and therapeutic applications. Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

30

Detailed Description

The drawings are first described.

Drawings

Fig. 1 is a schematic representation of the *in vitro* selection strategy used to isolate DNA molecules

- 11 -

having DNA ligase activity. Each molecule in the single stranded DNA (ssDNA) pool contained 116 random bases flanked by constant regions having sequences complementary to the PCR primers 5'-

5 GGAACACTATCCGACTGGCACC-3' (SEQ ID NO: 29) and 5'-biotin-CGGGATCCTAATGACCAAGG-3' (SEQ ID NO: 30). The pool was prepared by solid-phase phosphoramidite chemistry and amplified by PCR (Ellington et al., Nature 355:850-852, 1992) to yield approximately 32 copies of 3.5×10^{14}

10 different molecules. Single stranded DNA was prepared from the amplified pool as described by Bock et al. (Nature 355:564-566, 1992). The activated substrate (5'-biotin-AAGCATCTAAGCATCTCAAGC-p-Im (SEQ ID NO: 31)) contained a 5'-biotin group and a 3'-phosphorimidazolid

15 (Chu et al., Nucleic Acids Res. 14:5591-5603, 1986). Eight copies of the DNA pool (0.5 μ M) were incubated in selection buffer (30 mM Hepes, pH 7.4, 600 mM KCl, 50 mM $MgCl_2$, 1 mM $ZnCl_2$) with 1 μ M activated substrate and 1 μ M of an external template (5'-CGGATAGTGTTCGCTTGAGATGCTT-3'

20 (SEQ ID NO: 32)) complementary to the 5' end of the pool and the 3' end of the activated substrate. After a two hour incubation, the reaction was stopped by addition of EDTA. 0.5% ligated product was present after 24 hr. No product formation was observed in the absence of the

25 external template. At cycle 7, pool activity was independent of the external template, indicating that the remaining pool molecules were using an internal substrate binding site. In cycles 8 and 9, no external template was added, and the reaction time was decreased to 2 and

30 0.5 minutes, respectively, in order to increase selection stringency. To isolate ligated molecules, the reacted pool was passed through a streptavidin agarose affinity column (Pierce, Rockford, IL), unligated pool was washed off the column under denaturing conditions (3 M urea

35 followed by 150 mM NaOH, 40 column volumes each), and the

- 12 -

ligated pool was specifically eluted with excess free biotin (Wilson et al., Nature, in press, 1995). To select for substrate ligation to the 5'-hydroxyl of the pool molecules, isolated DNA was selectively PCR amplified (in cycles 6-9 only) with a first primer corresponding to the substrate sequence and a second primer complementary to the 3' constant region of the pool, and gel purified. This pool was then subjected to nested PCR with the first set of primers, gel purified, and re-amplified for ssDNA isolation (Bock et al., Nature 355:564-566, 1992). Nine cycles of selection-amplification were performed, after which the pool activity remained constant.

Fig. 2A is a denaturing acrylamide gel analysis of a time course of ligation reactions catalyzed by pool 9 ssDNA. Internally labeled pool 9 DNA (0.5 μ M) was incubated with activated substrate (1 μ M) in selection buffer for the indicated times. In a control reaction, the substrate was not activated (lane 5). DNAs were separated by electrophoresis in a 6% polyacrylamide/8 M urea gel. Radioactivity was detected using a Molecular Dynamics Phosphorimager.

Fig. 2B is a schematic representation of the sequences of clones isolated from pool 9 DNA. DNA from pool 9 was amplified by PCR and cloned into pT7Blue T-Vector (Novagen, Madison, WI). Each of the clones analyzed was sequenced in both directions using the standard dideoxy sequencing method. The 21 sequences (SEQ ID NOs: 1-21) shown in the figure share a consensus sequence consisting of two conserved domains (SEQ ID NOs: 22 and 23). Upper and lower case letters in the consensus indicate highly and moderately conserved positions, respectively. X and Z represent non-conserved, but complementary bases. The bolded T in domain I is present in 50% of the clones.

- 13 -

Fig. 3A is a schematic representation of the proposed secondary structure for the consensus sequence of the DNA molecules having DNA ligase activity isolated from pool 9 DNA. The 5' end of domain I and the 3' end of domain II base-pair with the 5' constant region of the pool (SEQ ID NO: 25 and the activated substrate, (SEQ ID NO: 24) respectively. The two complementary regions ("NNNN" of SEQ ID NO: 26 and "NNNN" of SEQ ID NO: 27) form a stem structure and bring the flanking domains into close proximity. Dotted lines indicate possible interactions between the bases at the ligation junction and the sequence between the two boxed sequences, TTT and AGA.

Fig. 3B is a schematic representation of a minimal DNA catalyst (SEQ ID NO: 28). Non-conserved regions in the DNA structure shown in Fig. 3A were deleted in order to generate a three-fragment complex in which the formation of a phosphodiester bond between the 3'-phosphorimidazolidine substrate S1 and the 5'-hydroxyl substrate S2 is catalyzed by the 47 nucleotide metalloenzyme E47.

Fig. 3C is a denaturing acrylamide gel analysis of a time course of ligation of activated substrate S1 and radiolabeled substrate S2 by the catalyst E47. No reaction was detectable when activated S1 (lanes 1 and 5) or E47 (lane 6) was absent.

Fig. 3D is a table showing the initial rates of ligation catalyzed by E47, E47-3T, E47-AGA, E47-hairpin, and pool 9 ssDNA. Activated substrate S1 (1 μ M) and radiolabeled S2 (0.5 μ M; S2 was 3'-end labeled using [α - 32 P]-cordycepin-5'-triphosphate (NEN Dupont, Boston, MA) and terminal transferase (Promega, Madison, WI)) were incubated with the different catalysts (0.75 μ M) at 25°C. Reaction conditions are as in Fig. 1, with the following changes: 30 mM Hepes, pH 7.2, and 4 mM ZnCl_2 . DNA was

- 14 -

separated by on a 12% polyacrylamide/8 M urea gel. K_{obs} values were determined by fitting fraction ligated vs. time to a linear equation using KaleidaGraph, and are the average of two independent experiments measured at less than 20% product formation. E47-3T and E47-AGA are E47 derivatives in which the conserved TTT and AGA sequences are deleted, respectively. E47-hairpin is an E47 derivative in which the hairpin has been replaced by 5'-CCATG-3'. The background reaction, containing an external template (see Fig. 1), was measured over a six hour incubation. No product was detected in the absence of the template, corresponding to a maximum rate of $2 \times 10^{-5} \text{ hr}^{-1}$.

Fig. 4A is a denaturing acrylamide gel analysis of an experiment showing the effect of Mg^{2+} , Zn^{2+} , and Cu^{2+} on catalysis. Reactions were incubated for 20 minutes at the indicated divalent metal ion concentrations. No reaction was detected in the absence of Zn^{2+} and Mg^{2+} (lane 2), or with only Mg^{2+} (lane 3). Mg^{2+} is not required for activity, and Zn^{2+} alone (lane 4) catalyzes the reaction with the same efficiency as Zn^{2+} and Mg^{2+} together. Cu^{2+} is the only divalent metal found that can substitute for Zn^{2+} (lane 5); it does not require Mg^{2+} for activity. The rate of ligation is independent of monovalent metal ions. Potassium chloride can be substituted by lithium, sodium chloride, or cesium chloride, or removed with no significant effect on product formation.

Fig. 4B is a graph showing the effects of zinc (○) and copper (●) concentrations on product formation. The reaction incubation time was 7 minutes.

Fig. 4C is a graph showing $\log(K_{obs})$ versus pH. In the presence of $10 \mu\text{M CuCl}_2$, there is a linear correlation between the log of K_{obs} and pH, with a slope of 0.7 up to pH 6.8. At higher pH values, the activity

- 15 -

decreases linearly with a slope of -0.7. A slope close to +1 suggests that proton abstraction is involved in the rate determining step of the reaction, while a slope of -1 is indicative of proton donation (Fersht, Enzyme Structure and Mechanism (Freeman, New York, 1985)). The observed rate is independent of buffer concentration between 30-150 mM. A similar effect was observed with Zn^{2+} at 4 mM up to pH 7.4. At higher pH, the activity drops drastically, possibly due to the formation of insoluble metal oxides or hydroxides (Bailar, Jr. et al., Comprehensive Inorganic Chemistry (Pergamon Press Ltd., 1973)). The reaction conditions were as specified in the description of Fig. 3.

Isolation of DNA molecule having DNA ligase activity

Oligodeoxynucleotides can be non-enzymatically ligated on either single-stranded (Naylor et al., *Biochemistry* 5:2722-2728, 1966) or duplex (Luebke et al., *J. Am. Chem. Soc.* 111:8733-8735, 1989) DNA templates. We designed an *in vitro* selection strategy (Szostak, *Trends Biochem. Sci.* 17:89-93, 1992; Chapman et al., *Curr. Opin. Struct. Biol.* 4:618-622, 1994; Breaker et al., *Trends Biotechnol.* 12:268-275, 1994; Joyce, *Curr. Opin. Struct. Biol.* 4:331-336, 1994) in order to determine whether DNA sequences which catalyze DNA ligation more efficiently than non-enzymatic templating could be isolated from a large pool of random sequences (Fig. 1). Using this strategy, a small single-stranded DNA that is a Zn^{2+}/Cu^{2+} -dependent metalloenzyme was isolated. The enzyme catalyzes the formation of a new phosphodiester bond by the condensation of the 5'-hydroxyl group of one oligodeoxynucleotide and a 3'-phosphorimidazolidine group on another oligodeoxynucleotide, and shows multiple turnover ligation.

- 16 -

The details of the selection strategy are illustrated in Fig. 1. After nine cycles of selection and amplification, the DNA pool (pool 9) displayed efficient ligation activity (Fig. 2A). Incubation of pool 9 DNA with the activated substrate yields a ligated product with the correct molecular weight and the expected nucleotide sequence at the ligation junction. To analyze further the selected sequences, DNA from pool 9 was cloned and sequenced. The majority of the clones contain a common consensus sequence consisting of two small domains separated by a spacer region of variable length and sequence (Fig. 2B). The two small domains are embedded in entirely different flanking sequences, indicating that several independent sequences in the original pool were carried through the selection process. Inspection of the consensus sequence suggests a secondary structure that is more complex than a simple template, but nevertheless brings the 5'-hydroxyl group and the 3'-phosphorimidazolidine group into close proximity (Fig. 3A).

Based on the consensus sequence, a small 47 nt ssDNA catalyst (E47) was designed that ligates two separate DNA substrates, S1 and S2 (Fig. 3B). Incubation of radiolabeled S2 with activated substrate S1 and E47 catalyst results in the appearance of the expected ligated product (Fig. 3C). Product formation requires that all three components are present in the reaction. In addition, the 3'-phosphate group of S1 must be activated. E47 catalyzes the ligation reaction twice as fast as pool 9. Small deletions within E47 result in severe losses of catalytic efficiency (Fig. 3D), indicating that the central consensus sequence is necessary for catalysis. The initial rate of ligation of S1 and S2 by E47 is 3400-fold greater than the rate of the same reaction catalyzed by a simple complementary

- 17 -

template under the same conditions, and is at least 10^5 -fold faster than the untemplated background ligation (Fig. 3D). This rate enhancement is comparable to values obtained for ribozymes obtained by *in vitro* selection (Szostak, Trends Biochem. Sci. 17:89-93, 1992; Chapman et al., Curr. Opin. Struct. Biol. 4:618-622, 1994; Breaker et al., Trends Biotechnol. 12:268-275, 1994; Joyce, Curr. Opin. Struct. Biol. 4:331-336, 1994) and catalytic antibodies (Lerner et al., Science 252:659-667, 1991).

10 Since the catalyst is not consumed in the reaction, it was expected that E47 would be capable of catalyzing the ligation of several molar equivalents of substrates S1 and S2, provided that the ligated product is able to dissociate from the enzyme. At saturating
15 concentrations (140 μM) of both substrates and 1 μM E47, multiple turnover catalysis at a rate of 0.66 hr^{-1} at 25°C and 2.4 hr^{-1} at 35°C was observed (10 turnovers observed). At these temperatures, product release appears to be rate limiting, as a rapid initial burst of approximately one
20 equivalent of product formation was observed within the first 10 minutes of the reaction. The initial rate of ligation in this burst phase was directly proportional to the concentration of E47 over a 30-fold range, as expected for an enzyme at saturating substrate
25 concentration (Persht, Enzyme Structure and Mechanism (Freeman, New York, 1985)). A plot of K_{obs} vs. [E47] yielded a k_{cat} of 3.2 hr^{-1} (0.07 min^{-1}) at 25°C .

Because divalent metal ions play a crucial role in ribozymes (Pyle, Science 261:709-714, 1993) and many
30 protein enzymes (Karlin, Science 261:701-708, 1993), it was expected that the DNA catalyst would require either Mg^{2+} and/or Zn^{2+} for activity, as these ions were present in the selection buffer. Indeed, the ligation reaction is dependent on Zn^{2+} (Fig. 4A), but does not require Mg^{2+} .
35 All of the members of the Irving-Williams series (Ba^{2+} ,

- 18 -

Sr²⁺, Ca²⁺, Mg²⁺, Mn²⁺, Fe²⁺, Co²⁺, Ni²⁺, Cu²⁺, Zn²⁺), as well as Pb²⁺ and Cd²⁺, were tested at concentrations between 10 μ M and 10 mM, and it was found that only Cu²⁺ could substitute for Zn²⁺. The efficiency of the ligation reaction is highly dependent on the divalent metal ion concentration (Fig. 4B). Increasing concentrations of Zn²⁺ up to 4 mM enhanced activity, but at higher concentrations the activity dropped sharply, suggesting the existence of inhibitory metal binding sites. A similar concentration dependence was observed for copper, but at a 400-fold lower concentration. The metal ion specificity suggests the existence of one or more metal ion binding sites with stringent geometrical and/or size requirements.

To gain insight into the ligation mechanism, the pH-rate profile of the reaction under pre-steady-state (single turnover) conditions was determined (Fig. 4C). The bell shaped profile displayed with Cu²⁺ suggests that the rate limiting step of the ligation reaction depends in part on two ionizable groups, once acidic and one basic, raising the possibility of a general acid-base mechanism (Fersht, Enzyme Structure and Mechanism (Freeman, New York, 1985)) in which copper complexes are involved in proton transfer. Metal-ion hydroxides are thought to act as general bases in some ribozyme-mediated RNA cleavage reactions (Pyle, Science 261:709-714, 1993; Dahm et al., Biochemistry 32:13040-13045, 1993; Pan et al., Biochemistry 33:9561-9565, 1994). Other possibilities, such as pH-dependent folding effects, may also account for these observations (Kao et al., Proc. Natl. Acad. Sci. USA 77:3360-3364, 1980).

E47 and substrates S1 and S2 were modified so that ligation of the modified substrates by the modified enzyme results in formation of a ligated product having the sequence of the modified enzyme. The sequences of

- 19 -

three such enzymes (E), and their corresponding substrates (S1 and S2), are as follows:

- I. E: 5'-
 5 ACCTTCACCTTCTTTTCGCTAGACCTTCAAGCGGAAGGTGAAGGT
 CTAGCG-3' (SEQ ID NO: 33)
- S1: 5'-ACCTTCACCTTCTTTTCGCTAGACCTTCAAGC-3'
 (SEQ ID NO: 34)
- S2: 5'-GGAAGGTGAAGGTCTAGCG-3' (SEQ ID NO: 35)
- II. E: 5'-
 10 ACCTTCACCTTCTTTTCGCTAGACCTTCAAGCGGAAGGTGAAGGT
 CTA-3' (SEQ ID NO: 36)
- S1: 5'-ACCTTCACCTTCTTTTCGCTAGACCTTCAAGC-3'
 (SEQ ID NO: 34)
- S2: 5'-GGAAGGTGAAGGTCTA-3' (SEQ ID NO: 37)
- 15 III. E: 5'-CTTCACCTTCTTTTCGCTAGACCTTCAAGCGGAAGGTGAAGGT
 CTA-3' (SEQ ID NO: 38)
- S1: 5'-CTTCACCTTCTTTTCGCTAGACCTTCAAGC-3'
 (SEQ ID NO: 39)
- S2: 5'-GGAAGGTGAAGGTCTA-3' (SEQ ID NO: 37)
- 20 The differences between these enzymes and E47 are in (1)
 the stem formed between E47 and the 5'-hydroxyl-
 containing substrate S2, (2) the stem formed between E47
 and the activated substrate S1, (3) the intramolecular
 stem in E47, and (4) the loop in E47. The sequence of
 25 the presumed core of the ligation site was not changed.
 The modified enzymes differ from one another only in the
 number of base pairs between the enzyme and the
 substrates. The modified enzymes catalyze ligation of
 their respective substrates, which shows that the primary
 30 nucleotide sequences of at least some parts of the stem
 and loop structures depicted in Fig. 3B are not required
 for enzyme activity, and further that the unchanged
 regions of the enzyme are sufficient for maintenance of

- 20 -

ligase activity, in the presence of the stem structures defined by S^1-E^1 and S^4-E^6 .

- 21 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: The General Hospital Corporation
- (ii) TITLE OF INVENTION: CATALYTIC DNA
- (iii) NUMBER OF SEQUENCES: 39
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Fish & Richardson P.C.
 - (B) STREET: 225 Franklin Street,
 - (C) CITY: Boston
 - (D) STATE: MA
 - (E) COUNTRY: USA
 - (F) ZIP: 02110-2804
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/487,867
 - (B) FILING DATE: 07-JUN-1995
 - (C) CLASSIFICATION:
- (vii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Lech, Karen P.
 - (B) REGISTRATION NUMBER: 35,238
 - (C) REFERENCE/DOCKET NUMBER: 00786/273001
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (617) 542-5070
 - (B) TELEFAX: (617) 542-8906
 - (C) TELEX: 200154

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 115 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

| | |
|---|-----|
| TATGTGTCGA TTGTGTTCTT TCGCTAGACC ATGTGAGACT TATGCTTCGA ATTGTCGACT | 60 |
| TTTGTACTGT TTGCTTGGCC CGCTGCTGGT CGTGCATGGT GAGATGATTA CCTA | 115 |

(2) INFORMATION FOR SEQ ID NO:2:

- 22 -

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 115 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TATGTGTGGA TTGTGTTCTT TCGCTAGACC ATGTGGGACT TATGCTTCCA ATTGTCGAGT 60
TTTGGACTGT TTGCTTGGCT GGCTGGTGGC CGCGCATGCT GAGATGATTA TCCCT 115

(2) INFORMATION FOR SEQ ID NO:3:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TATGTGTGGA TTGTGTTCTT TCGCTAGACC ATGTGAGACT TATGCTTCCA ATTGTCGAGT 60
TTTGGACTGT TTGCTTGGCC GGCTGCTGCT CGCGCATGCT GAGATGATTA TCCCTA 116

(2) INFORMATION FOR SEQ ID NO:4:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TATGTGTGGA TTGTGTTCTT CCGCTAGACC ATGTGAGACT TATGCTTCCA ATTGTCGAGT 60
TTTGGACTGT TTGCTTGGCC GGCTGGTGGT CGCGCATGCT GAGATGATTA TTCCCTG 117

(2) INFORMATION FOR SEQ ID NO:5:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 23 -

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TATAGTCAGG CTGGTAGGGT TCTTTCGCAG AGTGGGATGT GTTTTGATTI GAACTTATTT 60
ATGAGGTCTG TTGAAGCCCA TTGCGACTGA GTGCTTGCTG CTGTACTT TCCCTT 116

(2) INFORMATION FOR SEQ ID NO:6:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TATAGTCAGG CTGGTAGGGT TCTTTCGCAG AGTGGGATGT GTTTTGATTI GAACTTATTT 60
ATGAGGTCTG TTGAAGCCCA TTGCGACTGA GTGCTTGCTG CTGTACTT TCCCAT 116

(2) INFORMATION FOR SEQ ID NO:7:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TATAGTCAGG CTGGTAGGGT TCTTTCGCAG AGTGGGATGT GTTTTGATTI GAACTTATTT 60
ATGAGGTCTG TTGAAGCCCA TTGCGACTGA GTGCTTGCGG CTGTACTT TCCCAT 116

(2) INFORMATION FOR SEQ ID NO:8:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:8:

- 24 -

TATAGTCAGG CTGGTAGGGT TCTTTCGAG AGTGGGATGT GTTTTGATTG GAACTTATTT 60
ATCAGGTCGG TTGAAGCTCA TTCGGACTGA GTGCTTGCTG CTGTTACTT TCCCAC 116

(2) INFORMATION FOR SEQ ID NO:9:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGTTTCGTTT TCGAAGGCCT GTTGGTCCTT GTGTTCTCTC GCAGACCACT TTTTCGTACA 60
CGGAAGTGGA TTAAGTGGTG AGTTGCTTTC TAGTATGCGC TTTGAGGTAT TCTATG 116

(2) INFORMATION FOR SEQ ID NO:10:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CGTTTCGATT TCGAAGGCCT GTTGGTCCTT GTGTTCTCTC GCAGACCACT TTTTCGTTCA 60
CGGAAGTGGA ATAAGTGGTG AGTTGCTTTC TAGTGTGCGC TTTGAGGTAT TCTATG 116

(2) INFORMATION FOR SEQ ID NO:11:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGTTTCGTTT TCGAAGGCCT GTTGGTCCTT GTGTTCTCTC GCAGACCACT TTTTCGTTCA 60
CGGAAGTGGA TTAAGTGGTG AGTTGCTTTC TAGTGTGCGC TTTGAGGAAT TCTATG 116

(2) INFORMATION FOR SEQ ID NO:12:

- 25 -

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGTCTTGCTG GGTTTTGCT CGGTATCGTT CTTTCGCTAG ACCTTTAAAT AATGGTGAGA 60
TGCTGTTTTT GAGGCTAGTA GCGCGGGATT GGGCGTTACC GTCGTTTGTG TTTTGA 116

(2) INFORMATION FOR SEQ ID NO:13:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 115 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CACGTACTTC TTGTAGACGT GTGGCTTTGA TAGGATGTGG TCTTCGCTA GAGTTAATTA 60
GCTGTGGACC CTTAAGGTGT CTTAACTGAG ATGCTTTCAT TTTGTCTTTC TGATT 115

(2) INFORMATION FOR SEQ ID NO:14:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGCGTGGCT AACTGGATAG TGGTCTCTCG CTAGACACCT GTGTGAGATT GTTAGAATGC 60
GGTCCATCTG CCTATTGCT AGTTAAGGCT TTATGCTGTT CCTCTGATCA CTTTCG 116

(2) INFORMATION FOR SEQ ID NO:15:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 115 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 26 -

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTTTTGTGT TTGACGAATA CGTGTTCCTT CGCAGACCTT GTGCATCTTT GTTGTCGCAA 60
GGTGAGATGC TTGTGTTGTT TCCTTTTCA TGTTCCTTG TCCTGTTTT TARAC 115

(2) INFORMATION FOR SEQ ID NO:16:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TTGTGGTTGT GACCGGTTAG CATAGTGTTA TTTCGAGAC CACATCACCG TATTTGGTG 60
AGTGGTGAGA TCCTGCTATT TTGTGGTGT GCACCGCTT AAATACTTCG AGGTTT 116

(2) INFORMATION FOR SEQ ID NO:17:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTTGGTTTCG CAGTTGGTGT GTTCGTTGCG AGACCCTTTC GGTGAGATTG CTTTTCGGCC 60
TTTGACTGAT CCTGCCTTGT GGTATTGTTG TGCATGTGAT AGCTTGTCT GTCAT 116

(2) INFORMATION FOR SEQ ID NO:18:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:18:

- 27 -

TGGGGATCGC GGTATTAGTG TGTGCGTACT TTGGCTGACG GTGGCCGTCG TGGTATGTCT 60
GTTCTGTCCG ATGATCCAAT CTTCCCGGTT GGATGAGATG CTTGATTATG CTTA 114

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTTCTTGGGC TTAAGCTCGG TTATTGTTCT TTCGCTAGAT CCATGTCTAT ATTATGGTTG 60
GGCCGACTGG TTFTTTACTT ATACTATTGT TTTGTGGCG TGGATGAGAT GCTGTTT 117

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 116 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TCAGGTGTTT TTGTTTTTCT GAGCAGGGAG TCGGTGTGTT CTTTCGCAGA CACGAGTTTT 60
TTGTGTGAGA TTGCTTAGTG TTCTTTGTTT AATCACTAGA TTTCTTGATG GGTGTG 116

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 115 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTCGGTTCAT GTTGTTCCTT CGCCAGATGA TCGCGGCGTT TTAGTTTACG TCACTCGACG 60
TATTTTCTAC GGGGTTTAGG CTTTGTGGAT CATGAGTTGC TTAGATTGAT TTTT 115

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:

- 28 -

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CGGATAGTGT TCTTTCGCTA GANNNNN

27

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

NNNNNTGAGA TGCTT

15

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAGCATCTCA AGC

13

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

- 29 -

GGAACACTAT CCG

13

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CGGATAGTGT TCTTTGCTA GANNNN

26

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

HNNNTGAGAT GCTT

14

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 47 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CGGATAGTGT TCTTTGCTA GACCATGTGA CGCATGGTGA GATGCTT

47

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 30 -

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGAACACTAT CCGACTGGCA CC

22

(2) INFORMATION FOR SEQ ID NO:30:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CGGATCCTA ATGACCAAGG

20

(2) INFORMATION FOR SEQ ID NO:31:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AAGCATCTAA GCATCTCAAG C

21

(2) INFORMATION FOR SEQ ID NO:32:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CGGATAGTGT TCCGCTTGAG ATCCTT

26

(2) INFORMATION FOR SEQ ID NO:33:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

- 31 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACCTTCACCT TCTTCGCTA GACCTTCAAG CGGAAGGTGA AGGTCTAGCG

50

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ACCTTCACCT TCTTCGCTA GACCTTCAAG C

31

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGAAGGTGAA GGTCTAGCG

19

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 47 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ACCTTCACCT TCTTCGCTA GACCTTCAAG CGGAAGGTGA AGGTCTA

47

(2) INFORMATION FOR SEQ ID NO:37:

- 32 -

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CGAAGGTGAA GGTCTA

16

(2) INFORMATION FOR SEQ ID NO:38:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CTTCACCTTC TTTCGCTAGA CCTTCAAGCG GAAGGTGAAG GTCTA

45

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CTTCACCTTC TTTCGCTAGA CCTTCAAGC

29

What is claimed is:

- 33 -

CLAIMS

1. A method for obtaining a nucleic acid molecule having ligase activity, said method comprising the steps of:
 - 5 a) providing a population of candidate nucleic acid molecules, each having a region of random sequence;
 - b) contacting said population with:
 - (i) a substrate nucleic acid molecule; and
 - (ii) an external template complementary to a
 - 10 portion of the 3' region of said substrate nucleic acid molecule and a portion of the 5' region of each of the candidate nucleic acid molecules in said population, wherein binding of said external template to said substrate nucleic acid molecule
 - 15 and a candidate nucleic acid molecule from said population juxtaposes said 3' and 5' regions, and the terminal nucleotide of either said 3' or said 5' region contains an activated group;
 - c) isolating a subpopulation of nucleic acid
 - 20 molecules having ligase activity from said population;
 - d) amplifying said subpopulation *in vitro*;
 - e) optionally repeating steps b-d for said amplified subpopulation; and
 - f) isolating said nucleic acid molecule having
 - 25 ligase activity from said amplified subpopulation.
2. The method of claim 1, wherein said optional repeating of steps b-d is carried out in the absence of said external template.
3. The method of claim 1, wherein said nucleic
- 30 acid molecule having ligase activity or said substrate nucleic acid molecule is DNA.

- 34 -

4. The method of claim 1, wherein the 5' terminal nucleotide of said substrate nucleic acid contains a biotin moiety.

5 5. The method of claim 1, wherein said activated group is a 3'-phosphorimidazolidine on the 3' terminal nucleotide of said substrate.

6. A method for obtaining a DNA molecule having ligase activity, said method comprising the steps of:

- 10 a) providing a population of candidate DNA molecules, each having a region of random sequence;
- b) contacting said population with a substrate nucleic acid molecule;
- c) isolating a subpopulation of DNA molecules having ligase activity from said population;
- 15 d) amplifying said subpopulation *in vitro*;
- e) optionally repeating steps b-d for said amplified subpopulation; and
- f) isolating said DNA molecule having ligase activity from said amplified subpopulation.

20 7. The method of claim 6, wherein said substrate nucleic acid molecule is DNA.

8. The method of claim 6, wherein the 5' terminal nucleotide of said substrate nucleic acid contains a biotin moiety.

25 9. The method of claim 6, wherein said activated group is a 3'-phosphorimidazolidine on the 3' terminal nucleotide of said substrate.

10. A DNA molecule capable of acting as a catalyst.

- 35 -

11. A DNA molecule capable of acting as a catalyst on a nucleic acid substrate, said catalysis not requiring the presence of a ribonucleotide in said nucleic acid substrate.

5 12. A nucleic acid molecule having ligase activity.

13. The nucleic acid molecule of claim 12, wherein said nucleic acid molecule is DNA.

10 14. The nucleic acid molecule of claim 12, wherein said ligase activity is DNA ligase activity.

15 15. A nucleic acid molecule capable of ligating a first substrate nucleic acid to a second substrate nucleic acid, wherein the rate of said ligating is greater than the rate of ligating said substrate nucleic acids by templating under the same reaction conditions.

- 36 -

16. A catalytic DNA molecule capable of ligating a first substrate nucleic acid to a second substrate nucleic acid, said first substrate nucleic acid comprising the sequence 3'-S¹-S²-5', said second substrate
5 nucleic acid comprising the sequence 3'-S³-S⁴-5', and said catalytic DNA molecule comprising the sequence 5'-E¹-TTT-E²-AGA-E³-E⁴-E⁵-E⁶-3', wherein

S¹ comprises at least two nucleotides positioned adjacent to the 3' end of S², said S¹ nucleotides being
10 complementary to an equivalent number of nucleotides in E¹ that are positioned adjacent to the 5' end of said TTT;

S² comprises one - three nucleotides, S³ comprises one - six nucleotides, and the 5' terminal nucleotide of
15 S² and the 3' terminal nucleotide of S³ alternatively contain an activated group or a hydroxyl group;

S⁴ comprises at least two nucleotides positioned adjacent to the 5' end of S³, said S⁴ nucleotides being
complementary to an equivalent number of nucleotides in
20 E⁶ that are positioned adjacent to the 3' end of E⁵;

E¹ comprises at least two nucleotides positioned adjacent to the 5' end of said TTT, said E¹ nucleotides being complementary to an equivalent number of
nucleotides in S¹ that are positioned adjacent to the 3'
25 end of S²;

E² comprises zero - twelve nucleotides;

E³ comprises at least two nucleotides positioned adjacent to the 3' end of said AGA, said E³ nucleotides being complementary to an equivalent number of
30 nucleotides in E⁵ that are positioned adjacent to the 5' end of E⁶;

E⁴ comprises 3-200 nucleotides;

E⁵ comprises at least two nucleotides positioned adjacent to the 5' end of E⁶, said E⁵ nucleotides being
35 complementary to an equivalent number of nucleotides in

- 37 -

E³ that are positioned adjacent to the 3' end of said AGA; and

E⁶ comprises at least two nucleotides positioned adjacent to the 3' end of E⁵, said E⁶ nucleotides being
5 complementary to an equivalent number of nucleotides in S⁴ that are positioned adjacent to the 5' end of S³.

17. The catalytic DNA molecule of claim 16, wherein E² comprises three - four nucleotides.

18. The catalytic DNA molecule of claim 17,
10 wherein the 5' most nucleotide of S² is complementary to the 5' most nucleotide of E²; the 3' most nucleotide of S³ is complementary to the second 5' most nucleotide of E²; and the second 3' most nucleotide of S³ is complementary to the third 5' most nucleotide of E².

15 19. The catalytic DNA molecule of claim 18, wherein E² comprises four nucleotides, and the third 3' most nucleotide of S³ is complementary to the fourth 5' most nucleotide of E².

20 20. The catalytic DNA molecule of claim 16, wherein
a) S² comprises one nucleotide;
b) S³ comprises three nucleotides;
c) E⁴ comprises five nucleotides; or
d) E⁵ and E³ each comprise five nucleotides.

25 21. A method of ligating a first nucleic acid molecule to a second nucleic acid molecule, said method comprising contacting said first and said second nucleic acid molecules with a nucleic acid molecule having ligase activity.

- 38 -

22. The method of claim 21, wherein said nucleic acid molecule having ligase activity is DNA.

23. The method of claim 21, wherein said ligase activity is DNA ligase activity.

5 24. A nucleic acid molecule having ligase activity obtained by the steps of:

- a) providing a population of candidate nucleic acid molecules, each having a region of random sequence;
- b) contacting said population with:
 - 10 (i) a substrate nucleic acid molecule; and
 - (ii) an external template complementary to a portion of the 3' region of said substrate nucleic acid molecule and a portion of the 5' region of each of the candidate nucleic acid molecules from
 - 15 said population, wherein binding of said external template to said substrate nucleic acid molecule and a candidate nucleic acid molecule in said population juxtaposes said 3' and 5' regions, and the terminal nucleotide of either said 3' or said
 - 20 5' region contains an activated group;
- c) isolating a subpopulation of nucleic acid molecules having ligase activity from said population;
- d) amplifying said subpopulation *in vitro*;
- e) optionally repeating steps b-d for said
- 25 amplified subpopulation; and
- f) isolating said nucleic acid molecule having ligase activity from said amplified subpopulation.

 25. The nucleic acid of claim 24, wherein said optional repeating of steps b-d is carried out in the

30 absence of said external template.

- 39 -

26. The nucleic acid molecule having ligase activity of claim 24, wherein said nucleic acid molecule having ligase activity is DNA.

27. The nucleic acid molecule having ligase activity of claim 24, wherein

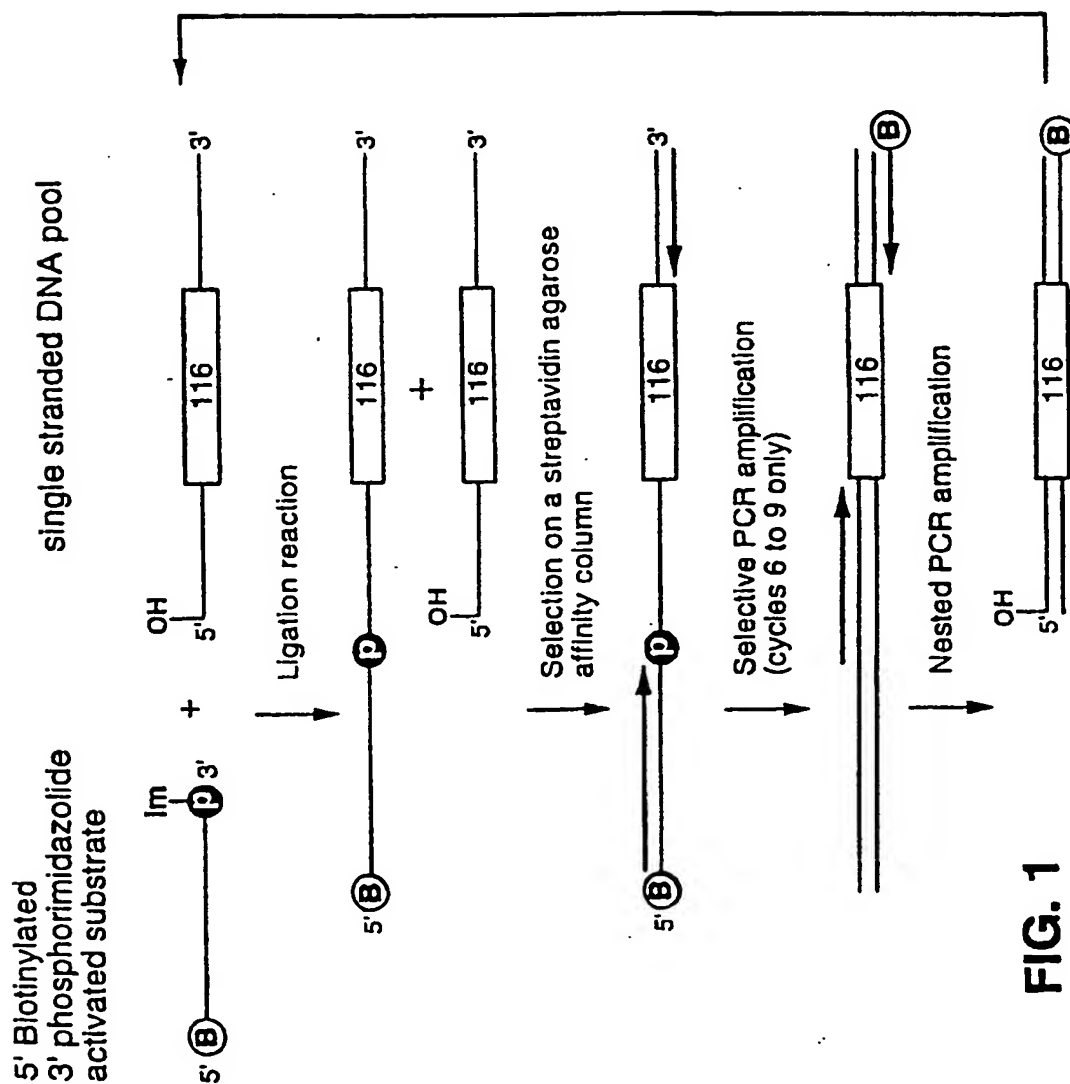
- a) the 5' terminal nucleotide of said substrate nucleic acid contains a biotin moiety; or
- b) said activated group is a 3'-phosphorimidazolidine on the 3' terminal nucleotide of said substrate.

28. A DNA molecule having ligase activity obtained by the steps of:

- a) providing a population of candidate DNA molecules, each having a region of random sequence;
- b) contacting said population with a substrate nucleic acid molecule;
- c) isolating a subpopulation of DNA molecules having ligase activity from said population;
- d) amplifying said subpopulation *in vitro*;
- e) optionally repeating steps b-d for said amplified subpopulation; and
- f) isolating said DNA molecule having ligase activity from said amplified subpopulation.

29. The DNA molecule having ligase activity of claim 28, wherein

- a) the 5' terminal nucleotide of said substrate nucleic acid contains a biotin moiety; or
- b) said activated group is a 3'-phosphorimidazolidine on the 3' terminal nucleotide of said substrate.



2 / 8

1 2 3 4 5

Ligated pool 9

Pool 9

reaction
time (min)

0 10 20 40 40

FIG. 2A

| | 1 | 2 | 3 | 4 | 5 | 6 |
|----------------------------|----|---|----|----|----|----|
| Activated S1 1 μ M | | + | + | + | | + |
| S2 0.5 μ M | + | + | + | + | + | + |
| E47 0.75 μ M | + | + | + | + | + | |
| Non-activated S1 1 μ M | | | | | + | |
| time min. | 30 | 5 | 10 | 30 | 30 | 30 |

Ligated S2

FIG. 3C

S2

3 / 8

| | | | | | | | |
|-------|--------------|--|------------------------------|-----------|----|----|----|
| | 1 | 10 | 20 | 30 | 40 | 50 | 60 |
| Seq01 | TATGT | GTCCGATTGTGTTCTTTCCGCTAGAGGATG | GTGAGACTTATGCTTCGAATTGTCGAGT | | | | |
| Seq02 | TATGT | GTCCGATTGTGTTCTTTCCGCTAGAGGATG | GTGGACTTATGCTTCGAATTGTCGAGT | | | | |
| Seq03 | TATGT | GTCCGATTGTGTTCTTTCCGCTAGAGGATG | GTGAGACTTATGCTTCGAATTGTCGAGT | | | | |
| Seq04 | TATGT | GTCCGATTGTGTTCTTTCCGCTAGAGGATG | GTGAGACTTATGCTTCGAATTGTCGAGT | | | | |
| Seq05 | TATAGTCAGGCT | GGTAGGGTTCTTTCCGAGAGTG | CGATGTGTTTTGATTGAACCTATTT | | | | |
| Seq06 | TATAGTCAGGCT | GGTAGGGTTCTTTCCGAGAGTG | CGATGTGTTTTGATTGAACCTATTT | | | | |
| Seq07 | TATAGTCAGGCT | GGTAGGGTTCTTTCCGAGAGTG | CGATGTGTTTTGATTGAACCTATTT | | | | |
| Seq08 | TATAGTCAGGCT | GGTAGGGTTCTTTCCGAGAGTG | CGATGTGTTTTGATTGAACCTATTT | | | | |
| Seq09 | CGTTTCGTTT | TGGAAGGCCCTGTTGGTCC | TTGTGTTCTCTCCGAGAGGACTTT | TTTCGTACA | | | |
| Seq10 | CGTTTCGATT | TGGAAGGCCCTGTTGGTCC | TTGTGTTCTCTCCGAGAGGACTTT | TTTCGTACA | | | |
| Seq11 | CGTTTCGTTT | TGGAAGGCCCTGTTGGTCC | TTGTGTTCTCTCCGAGAGGACTTT | TTTCGTACA | | | |
| Seq12 | CGTCTGCTGCGT | TTTGGCTCGTATCGTTCTTTCCGCTAGAGGATG | TTTAAATATGGTGAGA | | | | |
| Seq13 | CACGTACTTCT | TGTAGACGTGTGGCTTTGATAAGATGTTGGTCTTTCCGCTAGAGTTATTA | | | | | |
| Seq14 | GAGCCTGCTAA | CTGGATAGTGGTCTCTCCGCTAGAGGATG | CTGTGTCAGATTGTTAGAATGC | | | | |
| Seq15 | GTTTTTGTGTT | TGAAGAATACGTGTTCTTTCCGAGAGGTTT | GTGCATCTTTGTTGTCCGAA | | | | |
| Seq16 | TTGTGGTTGTG | ACCGGTTAGGATAGTGTTAATTTCCGAGAGGACTTT | CACCGTATTTTGGTG | | | | |
| Seq17 | TTTGGTTTCGC | AGTTGTTGTTCTCTCCGAGAGGCTTT | GGGTGAGATTGCTTTTGGCG | | | | |
| Seq18 | TGGGGATCGCG | GATTAGTGTCTGCGTACTTTGGCTGACGGTGGCCGCTG | GTGATGTTT | | | | |
| Seq19 | TTTCTTGGGCT | TAAGCTCGGTTATTTGTTCTTTCCGCTAGAGGATG | CTCTATATTATGTTG | | | | |
| Seq20 | TCAGGTGTTT | TGTTTTTCTGAGCAGGGAGTCGGTGTGTTCTTTCCGAGAGGAGT | TTT | | | | |
| Seq21 | GTCCGTTCA | TGTTGTTCTTTCCGAGAGGATG | CGCGGTTTTAGTTTACGTCACCTCGACG | | | | |

| | | | | | |
|----------------|---|----------------------------|---------------------|-----|-----|
| | 70 | 80 | 90 | 100 | 110 |
| TTTTGACTGTT | TGCTTGGCCGGCTGGTGGTCTG | CTATGGTGAGATGATT | ACCCTA | | |
| TTTTGACTGTT | TGCTTGGCTGGCTGGTGGCCGGC | CTATGGTGAGATGATT | TATCCCT | | |
| TTTTGACTGTT | TGCTTGGCCGGCTGGTGGTCTGCG | CTATGGTGAGATGATT | TATCCCTA | | |
| TTTTGACTGTT | TGCTTGGCCGGCTGGTGGTCTGCG | CTATGGTGAGATGATT | TATCCCTG | | |
| ATGAGGTCTG | TGAAGCCCATTCG | CACTGAGTGGCTT | GCTGCTTGTACTTTCCCTT | | |
| ATGAGGTCTG | TGAAGCCCATTCG | CACTGAGTGGCTT | GCTGCTTGTACTTTCCCAT | | |
| ATGAGGTCTG | TGAAGCCCATTCG | CACTGAGTGGCTT | GCGGCTTGTACTTTCCCAT | | |
| ATGAGGTCTG | TGAAGCTCATTG | CACTGAGTGGCTT | GCTGCTTGTACTTTCCAC | | |
| CGGAAGTGGAT | TAGTGGTGAGTTGCTT | CTAGTATGCGCTTTGAGCTATTCATG | | | |
| CGGAAGTGGAT | TAGTGGTGAGTTGCTT | CTAGTGTGCGCTTTGAGGATTCATG | | | |
| CGGAAGTGGAT | TAGTGGTGAGTTGCTT | CTAGTGTGCGCTTTGAGGAATTCATG | | | |
| TCCTCTTTT | TGAGGCTAGTAGCGCGGATTGGGCGTTACCGT | CGTTTGTCTTTCTGA | | | |
| GCTGTGGACCC | TAAAGGTGTCTT | CACTGAGATGCTT | CAATTTTGTCTTTCTGATT | | |
| GGTCCATCTG | CCTATTTGGTAGTTAAGGGTTTATGCTGTTTCTCTGATCACTTTTCG | | | | |
| GGTGAGATGCTT | GTGTTGTTGCTTTTTCATGTTTGTCTTGTCTTGTCTTTTAAAC | | | | |
| AGTGGTGAGATGCT | GCTATTTTGTGGTGTGACCCGCTTAAATACTTTCGAGGTTT | | | | |
| TTTGAGTGATC | CTGCTTGTGGTATTGTTGTGCATGTGATAGCTTGTCTGCTCAT | | | | |
| GTTCGTGCGATG | ATGCTTCTCCCGTTGQNTGAGATGCTT | GATTATGCTTA | | | |
| GGCCGACTGG | TTTTTACTTACTATTGTTTTTGTGCGGTGQNTGAGATGCTT | TTTT | | | |
| TTTGTGTGAGAT | TGCTTAGTGTCTTTGTTCAATCACTAGATTTCTTGATGGGTGTG | | | | |
| TATTTTCTAC | GGGTTTAGGCTTTT | CTGATCAATGAGTTGCTT | GATTGATTTTTT | | |

Consensus: 5' - (5-58) -

— cggataGTGTTCTTTTCGCTAGAGGATGCTT — (2-60) -

← Domain I →

— ZZZZTGAGATgctt — (4-62) - 3'

← Domain II →

FIG. 2B

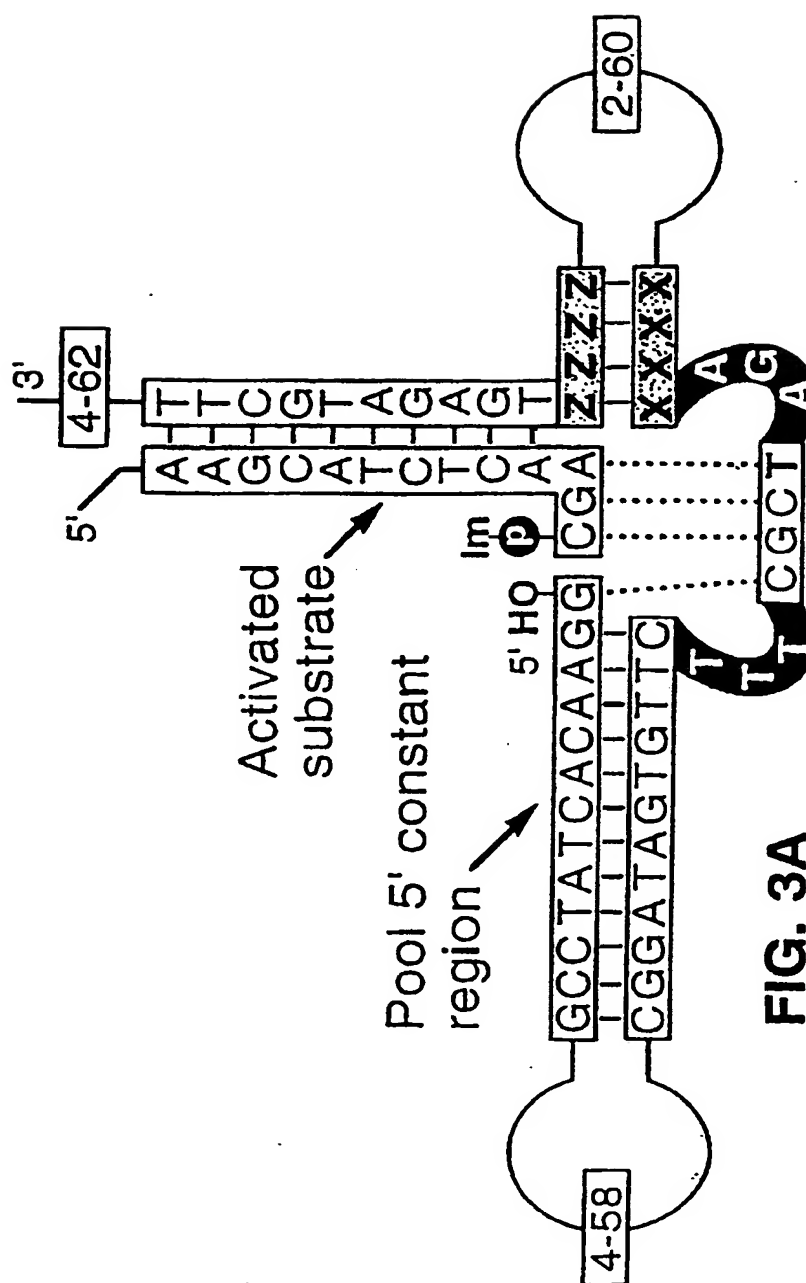
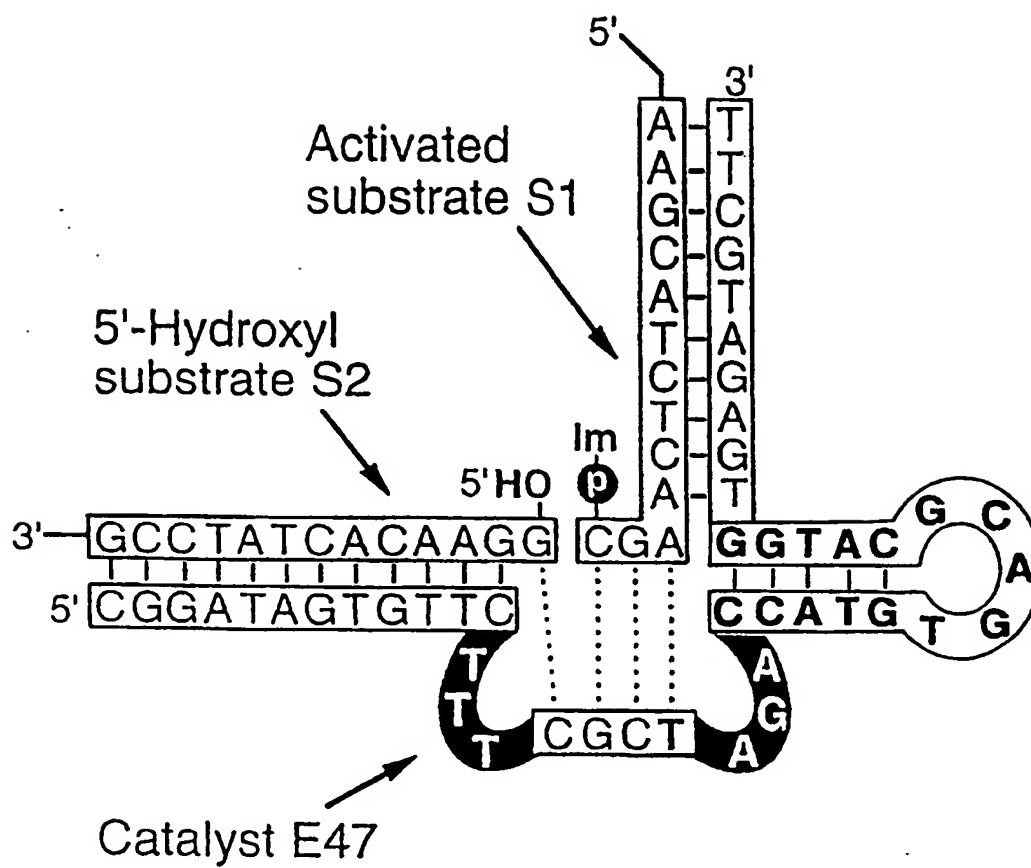
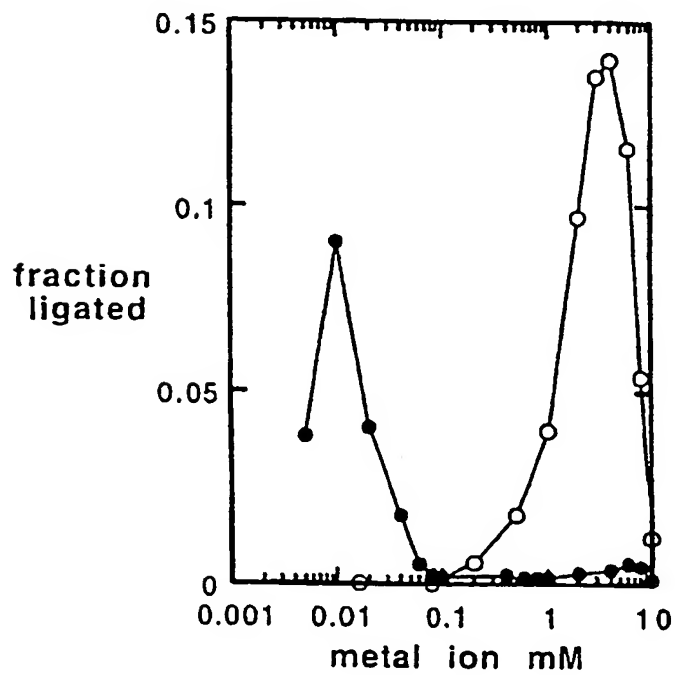


FIG. 3A

**FIG. 3B**

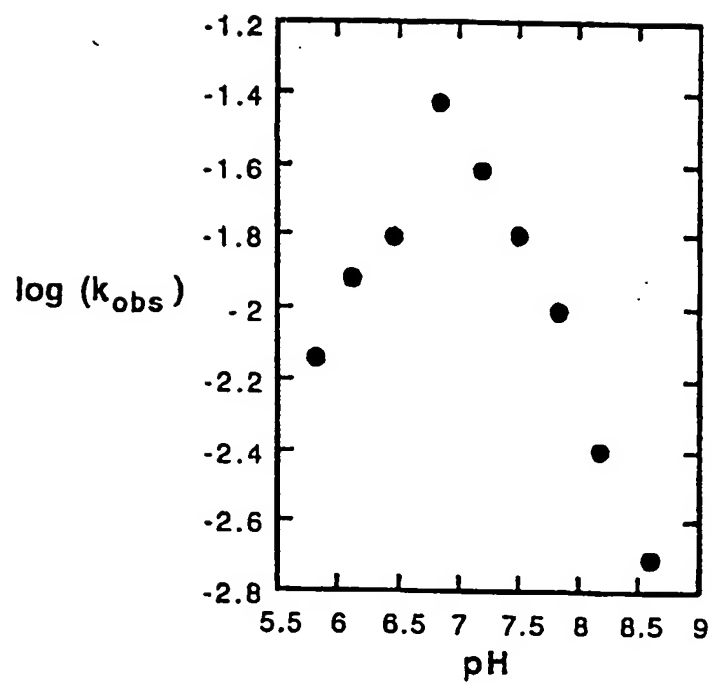
| | $k_{\text{obs}} \text{ hr}^{-1}$ |
|------------------|----------------------------------|
| E47 | 3.4 |
| E47-3T | < 0.01 |
| E47-AGA | < 0.01 |
| E47-hairpin | 0.41 |
| Pool 9 | 1.7 |
| templated bkgrd. | 0.0011 |
| background | < 2×10^{-5} |

FIG. 3D**FIG. 4B**

| | 1 | 2 | 3 | 4 | 5 |
|------------------------|---|---|---|---|---|
| Mg ²⁺ 50 mM | + | | + | | + |
| Zn ²⁺ 4 mM | + | | | + | |
| Cu ²⁺ 10 μM | | | | | + |



FIG. 4A

**FIG. 4C**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09358

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/04; C12Q 1/68; C12P 19/34

US CL :435/6, 91.2; 536/23.1, 25.4

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2; 536/23.1, 25.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS

search terms: deoxyribozyme, dna enzyme, catalytic dna, ribozyme, ligase

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|--------------|--|------------------------------|
| X -- A | YANG et al. Minimum ribonucleotide requirement for catalysis by the RNA hammerhead domain. Biochemistry. 1992, Vol. 31, pages 5005-5009, especially page 5006. | 10 ---- 1-9 and 11-29 |
| X -- A | CHARTRAND et al. An oligodeoxyribonucleotide with catalytic properties. Proc. RNA Soc. 1994, Vol. 77, page 5. | 10 ----- 1-9 and 11-29 |
| X -- A | BREAKER et al. A DNA enzyme that cleaves RNA. Chem. Biol. December 1994, Vol. 1, pages 223-229 | 10 ---- 1-9 and 11-29 |

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

| | | |
|---|-----|--|
| * Special categories of cited documents: | * T | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| * A* document defining the general state of the art which is not considered to be of particular relevance | * X | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| * E* earlier document published on or after the international filing date | * Y | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| * L* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | * G | document member of the same patent family |
| * O* document referring to an oral disclosure, use, exhibition or other means | | |
| * P* document published prior to the international filing date but later than the priority date claimed | | |

Date of the actual completion of the international search

31 JULY 1996

Date of mailing of the international search report

05 SEP 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Authorized officer

KENNETH R. HORLICK

Facsimile No. (703) 305-3230

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09358

| C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|---|--|--|
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X — A | BARTEL et al. Isolation of new ribozymes from a large pool of random sequences. Science. 10 September 1993, Vol. 261, pages 1411-1418, especially page 1412. | 12, 15, 21, 24-25, and 27 ----- 1-11, 13-14, 16-20, 22-23, 26, and 28-29 |